

基于扩展的 PCANet 的有遮挡人脸识别方法^①秦斌^{②*} 卢天宇* 李卫锋** 刘银伟* 朱娅妮^{③***} 李小薪*

(*浙江工业大学 计算机科学与技术学院 杭州 310023)

(**浙江华通云数据科技有限公司 杭州 310018)

(***)杭州电子科技大学 网络空间安全学院 杭州 310018)

摘要 针对有遮挡人脸识别问题,将现有的 CNN 模型与 PCANet 相结合,提出了扩展的 PCANet 模型。为了有效消除人脸图像中可能包含的遮挡信息造成的影响,通常需要充分利用网络的底层特征并构建尽可能丰富的特征。PCANet 的两个不足在于:由于正交性约束,各卷积层的滤波器高度相似,降低了滤波器响应的多样性;在进行模式图编码时,对特征图进行了二值化处理,并采用了跨度较大的编码方式,从而丢弃了过多的信息。为了使 PCANet 能够更好地适配于现有的 CNN 模型,在 PCANet 模型中引入了两个稠密连接:在各卷积层之间引入了稠密连接,以充分利用底层卷积层提取的特征,并尽可能降低卷积层之间的滤波器的相似性;在 PCANet 的模式图编码阶段引入了加权稠密编码,以充分利用卷积层输出的特征生成更多的模式图。这两种稠密连接或编码方案都会进一步提升 PCANet 最终输出的柱状图特征的维度,并生成更为丰富的特征。在受控环境和有真实遮挡的人脸数据集 (AR 人脸数据集)、非受控环境和有模拟遮挡的数据集 (LFW 和 CFP)、非受控环境和有真实遮挡的数据集 (MFR2 和 PKU-Masked-Face) 上的实验表明,所提出的扩展的 PCANet 模型能够有效处理实物遮挡和因光照引发的遮挡,也可以作为前沿方法的有效补充,提升前沿方法的遮挡鲁棒性。

关键词: 有遮挡人脸识别, PCANet, 稠密连接, 稠密编码, 滤波器多样性

0 引言

近年来,以卷积神经网络 (Convolutional Neural Networks, CNN) 为代表的深度神经网络 (Deep Neural Network, DNN) 在计算机视觉领域中获得了巨大的成功,优秀的深度学习模型在一些公开的数据集上的分类能力甚至超过了人类,比如:在 LFW 人脸数据库上的认证准确率^[1,2],在 ImageNet 上的图像分类准确率^[3,4]。然而,在实际情景中,待识别图像在“结构”或者“分布”上经常与训练图像存在较大的差异,这种差异会降低深度学习模型的识别率。在计算机视觉领域中,这种待识别图像与训练图像或库存图像有较大差异的现象也被称作为“协变量偏移 (Covariate Shift)”问题^[5-7]。当待识别图像与训练集图像存在遮挡、光照、分辨率等幅度较大的差异时,现有的 CNN 模型的识别准确率往往会大幅度下降。本文重点关注遮挡造成的人脸图像的“偏移”问题。

为了减轻遮挡造成的性能退化,前沿的研究工作给出了

两种解决方案:基于图像重建的方法^[8-11]和基于遮挡去除的方法^[12-14]。基于图像重建的方法首先修复被遮挡的图像区域,再对修复后的人脸图像进行识别。Zhao 等人^[8]首先提出了基于 LSTM 自动编码器对人脸遮挡区域进行重建的人脸识别模型。此类方法的主要挑战在于如何在较好重建遮挡区域的同时保留人脸图像的判别信息。另一类方法旨在去除遮挡的影响:检测并删除被遮挡破坏的特征,再利用剩余的干净特征进行识别。此类方法的主要挑战是如何检测到遮挡在人脸图像或人脸图像的特征图中的位置。Wan 等人^[12]通过向 CNN 模型的中间层添加一个 MaskNet 分支来解决这一问题,该分支通过学习掩码权重向量,将较低的权重分配给被遮挡损坏的特征图。Song 等人^[13]提出学习二值掩码矩阵,在网络的最后一个隐含层,去除被遮挡损坏的特征。为此,需要将人脸图像划分为 $K \times K$ 块,在训练阶段学习 K^2 个二值的遮挡掩码字典,得到不同的遮挡块和不同的字典原子之间的映射关系;在测试阶段,首先检测被遮挡的块,然后检索遮挡

① 本文受浙江省自然科学基金(LGF22F020027)、国家自然科学基金(62271448)等资助。

② 女,1978 年生,硕士,讲师;研究方向:图像处理与模式识别;E-mail: qe@zjut.edu.cn

③ 通信作者, Email: 10764341@qq.com

(收稿日期: 2023-09-17)

字典中对应的字典原子,并将其用于去除测试特征中的坏特征。这种两阶段的方法须依赖于一个用于遮挡检测的外部网络,并且 Song 等人^[13]提出的方法需要在训练阶段独立训练 K^2 个深度网络模型来学习遮挡字典,这使得训练效率低下且非常耗时。最近, Qiu 等人^[14]提出一种端到端的有遮挡人脸识别方法:在训练阶段,利用遮挡的空间连续性作为特征掩码学习的监督信号,学习掩码解码器;在测试阶段,掩码解码器会根据输入图像动态生成一个新的特征掩模,并使用该特征掩模清除特征图中被遮挡损坏的特征,从而达到鲁棒识别的目的。

上述方案都需要对遮挡信息进行明确的处理:或者对遮挡区域进行重建,或者去除遮挡的影响。这会导致两个问题: *i)* 需要花费较大的代价对“遮挡”本身进行建模; *ii)* 已有的在大规模数据集上训练的经典的 CNN 模型,如 VGGFace^[15]、ResNet^[16]、FaceNet^[2]、SphereFace^[17]等,很难直接被用于有遮挡人脸识别,造成了巨大的浪费。那么,是否存在一种方法既不需要对遮挡进行显式处理,也能很好地解决遮挡问题呢?已知的方法主要有两种:基于图像梯度方向 (Image Gradient Orientations, IGO) 域的鲁棒误差度量方法^[18, 19]和基于 FPH 框架^[20-28]的特征提取方法。基于 IGO 的鲁棒误差度量方法由 Tzimiropoulos 等^[18]提出,他们指出:两幅完全不同的图像在 IGO 域中的差值以非常高的置信度服从 $[0, 2\pi)$ 上的均匀分布,且总体差异之和几乎为零。利用这一现象,可以有效消除遮挡的影响。然而,正如 Li 等人^[19]所指出的,由于在实际中很难存在理想的与人脸图像“完全不同”的遮挡,在 IGO 域中的鲁棒误差度量并不能完全消除遮挡的影响,还需要辅之以遮挡检测或遮挡区域的重建。与鲁棒误差度量方法不同,以 LBP^[21]和 PCANet^[29]为代表的 FPH 方法通过提取局部统计特征而达到鲁棒识别的目的。具体的, FPH 方法主要包括特征图 (Feature-maps) 提取、模式图编码 (Pattern-maps) 和柱状图 (Histogram) 统计等三个步骤:特征图提取旨在去除干扰信号,提取判别性信息,有效的特征提取可以极大提升识别鲁棒性;模式图编码首先对特征图进行二值化处理,进一步抑制所提取的特征图中包含的噪声信息,并通过整型编码,重组二值化后的特征图,

以有效融合多个特征图之间的信息,形成鲁棒的模式图;柱状图统计旨在获得高维的统计特征:将输入的模式图分割成多个较小的局部块,再从不同的局部块提取柱状图特征,形成高维稀疏的统计特征¹,高维稀疏的统计特征可以在很大程度上“稀释”掉由遮挡等噪声所导致的坏特征,从而达到遮挡鲁棒性。特征图提取对于 FPH 方法至关重要,但现有的 FPH 方法通常只使用了较为浅层的特征提取方法,很难有效处理现实世界中人脸图像中广泛存在的各种变化,即很难适用于非受控环境下的人脸识别。

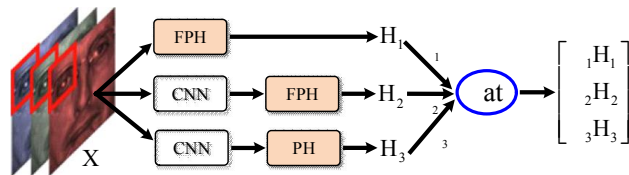


图1 扩展的 FPH (extended FPH, xFPH) 框架,其中, Cat 表示对特征图进行拼接的算子

为了使 FPH 方法能够更好的适用于非受控环境,也为了充分利用现有的 CNN 模型,使其具有遮挡鲁棒性,本文将 FPH 方法和现有的 CNN 模型^[2, 15-17]结合起来,充分利用二者的互补关系,构建遮挡鲁棒的人脸识别模型。具体的,如图1所示,将现有的 CNN 模型看作黑盒子,假定 CNN 模型的输出为其最后一个卷积层的输出,构建三个分支: FPH 分支、CNN+FPH 分支、CNN+PH 分支,并将它们输出的柱状图特征加权拼接起来。这里, PH 表示“模式图编码+柱状图统计”, CNN+PH 表示直接对 CNN 输出的特征图进行模式图编码,而不是使用 FPH 自身的特征提取方法。图1所示的特征提取方法可以看作是对 FPH 框架的扩展,本文称之为扩展的 FPH (extended FPH, xFPH) 框架。

FPH 框架包含一系列方法^[20-28],不同的 FPH 方法具有不同的特征提取方法和模式图编码方法,与 CNN 模型的适配程度是不同的,本文选取其中的 PCANet 模型,探讨其在 xFPH 框架下的遮挡鲁棒性。这里,选取 PCANet 是因为 PCANet 是 FPH 系列方法的优秀代表,具有很好的遮挡鲁棒性^[29-31],并且 PCANet 所使用的特征提取方法也是建立在卷积特征的基础上的,因此 PCANet 的模式图编码方法也能

¹ 一般情况下,所分割的局部块的尺寸越小,所获取的局部柱状图特征越稀疏,维度也越高。

更好地适用于卷积特征。

为了使 PCANet 能够更好地适配于 CNN 模型，更好地进行有遮挡人脸识别，需要对 PCANet 进行改进。PCANet 的特征的遮挡鲁棒性也是建立在高维、稀疏、局部统计性的基础上的。PCANet 只利用了最后一个卷积层输出的特征图来生成后续的模式图和柱状图特征。这一方面极大浪费了底层卷积层所输出的特征，不利于形成更高维度的特征；另一方面，正如 Wright 等^[9]所指出的，深层特征提取会将遮挡的影响扩散到更大的邻域范围，这就意味着，对于有遮挡人脸识别而言，网络的底层特征往往也是至关重要的，需要充分利用。因此，本文将稠密连接^[32]引入 PCANet 的特征提取阶段（详见 4.1 节），以充分利用底层特征，并形成更高维的特征。

另一方面，为了形成更好的遮挡鲁棒性，PCANet 在进行柱状图统计时，需要使用较小的滑动窗口（如 3×3 ）以减少潜在的遮挡块对邻域特征块的影响，这会导致所形成的柱状图特征非常稀疏。稀疏性对于减少坏特征的影响具有重要意义，但也不利于提取判别性特征——判别性往往建立在特征多样性^[33, 34]的基础上。为了平衡稀疏性和判别性，可以在同一个滑动窗口的更多个模式图中提取柱状图特征。PCANet 在进行模式编码时，采用了较大的跳跃步长，既未能充分利用各特征图之间的相关性，也导致所形成的模式图的个数较少。为了获取更多的模式图，并充分挖掘其中更具有遮挡鲁棒性的模式图，本文提出了加权稠密编码方法（详见 4.2 节）。

综上，本文将稠密连接和加权稠密编码引入 PCANet，提出了稠密 PCANet (Dense PCANet, DPCANet)，以尽可能地丰富 PCANet 所提取的特征，从而提升 PCANet 的遮挡鲁棒性；并将 DPCANet 用于 xFPH 框架（图 1），构建扩展的 PCANet 模型 (extended PCANet, xPCANet)，以使现有的 CNN 模型能更好地适用于有遮挡人脸识别。实验表明，xPCANet 能够更加有效的处理待识别图像中的遮挡、光照等变化，能够有效地提升现有 CNN 模型对有遮挡人脸图像的识别率。

1 问题的定义

本文重点考虑训练集图像与测试集图像发生严重“偏移”的有遮挡人脸识别问题，如图 2 所示，训练集由自然光照下采集的不含遮挡的人脸图像组成，而测试集由含有遮挡和不均匀光照的人脸图像组成。令 $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^{h \times w \times c}\}_{i=1}^N$ 表示 N 张训练集图像的集合， $\mathbf{Y} = \{\mathbf{y}_i \in \mathbb{R}^{h' \times w' \times c}\}_{i=1}^K$ 表示 K 张存在光照和遮挡变化的待识别图像的集合（测试集），其中， $c \in \{1, 3\}$ 表示训练/测试图像的通道（ $c = 1$ 表示灰度图像， $c = 3$ 表示 RGB 彩色图像）。

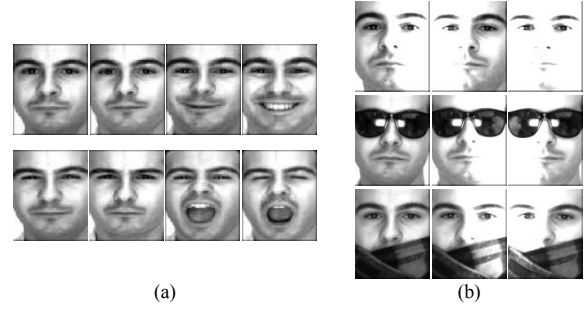


图 2 有遮挡人脸识别问题示例：(a) 训练集 \mathbf{X} 中样本示例：8 张不含遮挡和光照变化的人脸图像；(b) 测试集 \mathbf{Y} 中样本示例：9 张有遮挡和光照变化的人脸图像。

由于测试图像与训练图像在像素分布上存在较大差异，直接在像素域中进行特征匹配很容易导致错误识别（预测），例如，黑色围巾遮挡很容易匹配到黑色的胡子从而导致识别错误。本文探讨如何基于扩展的 PCANet (xPCANet) 有效补偿有遮挡人脸图像中丢失的判别信息和对抗因遮挡而引发的错误匹配，从而达到鲁棒识别的目的。xPCANet 提取的是柱状图特征，因此，使用基于卡方距离 $\chi^2(\cdot, \cdot)$ ^[29] 的最近邻方法进行分类。对于 $\forall \mathbf{y} \in \mathbf{Y}$ ，从 \mathbf{X} 中寻求满足如下优化式的 \mathbf{x} ：

$$\mathbf{x}^* = \underset{\mathbf{x} \in \mathbf{X}}{\operatorname{argmin}} \chi^2(f_{\text{xPCANet}}(\mathbf{x}), f_{\text{xPCANet}}(\mathbf{y})) \quad (1)$$

从而确定 \mathbf{y} 的类别： $\text{Id}(\mathbf{y}) = \text{Id}(\mathbf{x}^*)$ 。

2 PCANet 的主要缺陷

为了使 PCANet 能更好地适用于现有的 CNN 模型，需要探讨 PCANet 存在的主要缺陷。PCANet^[29] 采用了经典的 FPH 框架^[35]：特征图 (Feature Map) 提取-模式图 (Pattern Map) 编码-柱状图 (Histogram) 统计。与经典的 FPH 方法（如 LBP^[21]）不同的是：(1) PCANet 在特征图提取阶段采用了

深度卷积提取特征图；（2）PCANet 在模式图编码阶段采用了“立体式稀疏编码”。然而，PCANet 在这两个阶段所采用的处理策略均存在一定的缺陷：在特征提取阶段，PCANet 所采用的滤波器组存在较大的冗余性，详见 3.1 节；而在模式图编码阶段，PCANet 所采用的“立体式稀疏编码”未能充分利用 PCANet 的特征图信息，详见 3.2 节。

2.1 PCANet 滤波器的冗余性

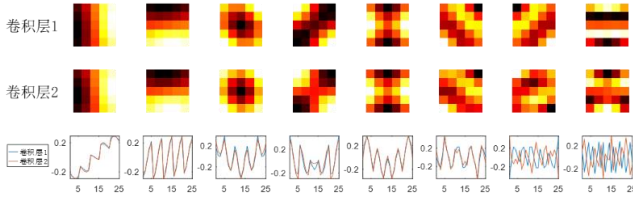


图 3 PCANet 所采用的两组 5×5 的滤波器。第 1 行和第 2 行：卷积层 1 和卷积层 2 的滤波器组的二维呈现；第 3 行：滤波器组的一维呈现（横坐标为滤波器的索引，纵坐标为对应的滤波器值）

PCANet 的一个主要不足在于其滤波器组中的滤波器缺乏多样性。如图 3 所示，PCANet 通常只采用 2 个卷积层的 16 个卷积核（每个卷积层 8 个卷积核），且这两组滤波器高度相似：既缺乏像 Gabor 滤波器那样丰富的尺度和方向，也缺乏像经典的 CNN 那样的庞大的参数规模；从图 3 第三行的一维展示来看，在 PCANet 所生成的两层 8 个滤波器中，除了最后两个滤波器的波峰波谷位置发生了变化外，其它 6 个滤波器几乎完全相同。

导致 PCANet 的滤波器组具有高度相似性的原因主要有两个：（1）随着卷积深度的增加，参与提取 PCA 滤波器的图像块也越来越多，这些图像块被不加区分地堆叠在一起，这就导致 PCANet 只是从少数变化显著的图像块中学习 PCA 滤波器，而这些变化显著的图像块在各卷积层之间都是非常相似的；（2）PCA 滤波器是在低维（如 $9 = 3 \times 3$ 或 $25 = 5 \times 5$ ）子空间中获取的，显然，子空间维度越低，图像块或特征块的数值动态范围就越小，所获取的 PCA 滤波器就具有越高的相似性和越差的结构性；即使在同一维度空间中，所获取的 PCA 滤波器也会随着特征值的衰减（也就是在相应方向上特征变化的衰减）而迅速丧失结构性，从图 3 中滤波器在一维呈现下的震荡情况也可以观察到这一

现象：越缺乏结构性的滤波器的震荡越不规则。因此，只是通过简单地增加卷积层或增加滤波器的个数并不能有效提升 PCA 滤波器的多样性。Chan 等^[29]的实验表明，超过 2 个卷积层的 PCANet 对于性能提升并没有太大作用。

为了提升 PCANet 滤波器的多样性，Low 等^[36]提出使用 Gabor 滤波器与 PCA 滤波器或 ICA 滤波器交叉卷积的方式来生成更为丰富的滤波器组。然而，由于 Gabor 滤波器是手工设计的，不能很好地刻画蕴藏于数据中的本质特征，因此，对于有遮挡人脸识别问题而言，同样会导致识别性能的不稳定性。

由于增加深度和借鉴已有的手工设计的滤波器都无法有效提升 PCANet 滤波器的多样性，那么是否可以直接提升特征图的多样性呢？受稠密连接思想^[32, 37]的影响，为了提升特征的多样性，可以采用“特征重用”策略，将 PCANet 的低层卷积特征与高层卷积特征联合使用，从而达到丰富特征图的目的，详见图 5 及 4.1 节。

2.2 PCANet 模式图的立体式稀疏编码

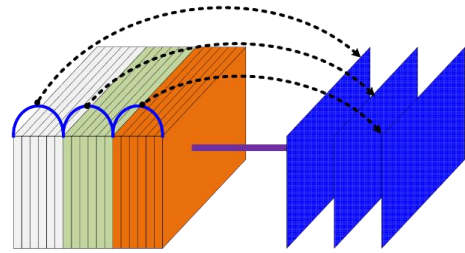


图 4 立体式稀疏编码模式

在特征提取阶段，PCANet 所采用的卷积与经典的 CNN 所使用的卷积不同：PCANet 采用了单通道卷积（Single-channel Convolution）或者“通道无关式卷积（Channel-Independent Convolution, CIC）”，而经典的 CNN 通常使用多通道卷积（Multi-channel Convolution）或者“通道相关式卷积（Channel-Dependent Convolution, CDC）”。也就是，PCANet 在特征提取阶段忽略了特征图在通道方向上的相关性，然而，通过使用 CIC 可以将各滤波器与各输入的特征图以相互组合的方式生成新的特征图，是以一种比较经济的方式（用尽可能少的滤波器）产生尽可能丰富的特征图。

这一“忽略通道方向上的相关性”的不足在模式图编码

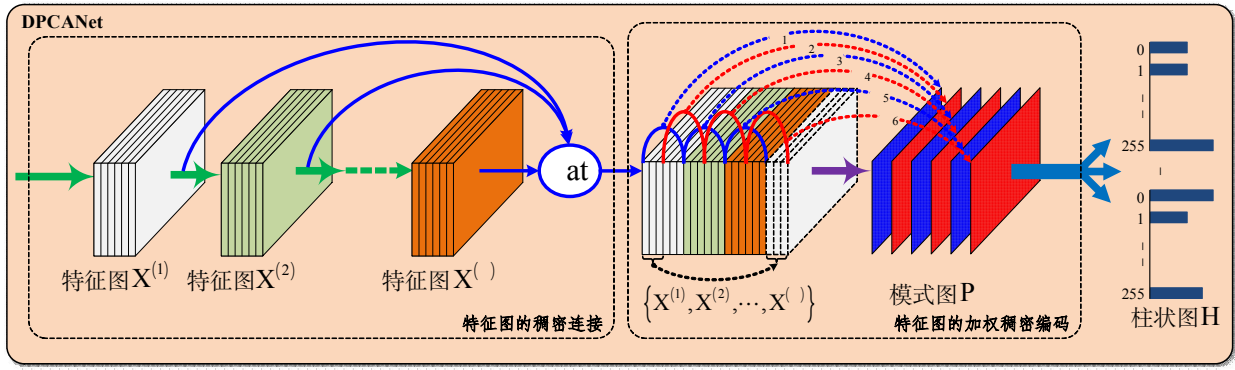


图5 稠密 PCANet 的网络结构图，其中， $\langle \mathbf{block}' \rangle$ $\text{xmlns}=\text{'http://www.w3.org/1998/Math/MathML'}$

阶段得到了纠正: PCANet 采用了不同于经典的 FPH 方法的平坦式编码方式 (如 LBP^[21]是使用单个特征图生成单个模式图的, 本文称之为“平坦式编码”), 在多个通道上进行模式图编码, 即采用如图 4 所示的“立体式稀疏编码 (Stereo Sparse Coding, SSC)”。

除了需要进行特征压缩外, PCANet 采用立体式编码而非平坦式编码的主要原因在于: PCANet 特征的二值化不需要与邻域特征点比较。首先, PCANet 已经采用了卷积对输入图像的每个像素与其周围像素点的局部相关性进行了计算, 在编码阶段则只须集中考虑各特征点在通道方向上的相关性; 其次, 正如 Li 等^[35]所指出的, PCANet 的卷积运算实际上是对特征进行方向分解, 在方向确定的情况下, 只需要关心特征在这一方向的正负即可, 因此, PCANet 使用单位阶跃函数 (Unit Step Function) 进行特征的二值化, 也就是每个特征点通过与 0 比较进行二值化。

尽管特征图在通道方向上的相关性在模式图编码阶段进行了一定程度的计算, 然而, 由于 PCANet 在进行模式图编码时对特征图进行了二值化处理, 特征图之间的相关性已经被大大削弱了。那么, 是否可以直接在特征图提取阶段进行通道方向上的相关性计算? 另一方面, 由于 PCANet 采用了“稀疏”的编码方式 (在通道方向上的编码步长默认设置为 8), 特征图所包含的信息被极大的压缩了, 对于有遮挡图像而言, 这样的压缩可能会导致很多判别信息的丢失, 本文将通过对特征图进行稠密编码 (详见图 5 及 4.2 节), 以提升特征图的利用率, 从而增强最终形成的柱状图特征的多样性。

3 扩展的 PCANet

本文所提出的扩展的 PCANet (xPCANet) 模型是将稠密 PCANet (Dense PCANet, DPCANet) 用于 xFPH (图 1) 所得到的。DPCANet 的网络结构如图 5 所示, 主要包括特征提取阶段的稠密连接和模式图编码阶段的加权稠密连接, 4.1 节和 4.2 节将分别介绍 DPCANet 的这两个阶段的稠密连接。DPCANet 在模式图编码阶段引入了权重参数 $\{\alpha_i\}$, xPCANet 所使用的 xFPH 框架也引入了 ω_1 、 ω_2 和 ω_3 等三个权重参数, 因此, xPCANet 中包含了两种不同类型的权重参数。4.3 节将进一步介绍这些权重参数的计算方法。

3.1 特征图的稠密连接

正如 Li 等^[35]所指出的, PCANet 的各卷积层在本质上是对输入图像进行了多尺度 (不同卷积层的感受野不同) 和多方向 (在不同的主方向上的投影) 分解。具体的, 由于使用了深层滤波, 更高层的卷积层是在更大的尺度或感受野²上获取特征图; 而 PCANet 所采用的 PCA 滤波在本质上是一种“方向滤波”^[35]: 输入特征被分解到特征变化最为显著的几个方向 (主方向)。如何获得并充分利用输入图像的多尺度多方向特征一直是特征提取的一个指导原则^[38-41], 然而, PCANet 丢弃了低层卷积层 (较小尺度) 的特征图, 只使用了最后一个卷积层的特征图。而对于有偏移图像的识别而言, 低层卷积层产生的特征图往往比高层卷积层产生的特征图更具有判别性, 这是因为高层卷积层由于具有较大的感受野,

² 假设当前卷积层的卷积核尺寸为 $k \times k$, 其感受野相对于上一个卷积层都增大了 $(k+1)/2$ 。

更容易将偏移所导致的局部坏特征在更大的范围内传播到邻域特征点。

为了充分利用 PCANet 的多尺度多方向特征, 将稠密连接^[32, 37]引入 PCANet 的基于 CIC 的卷积块, 如图 5 所示。这里, 稠密连接通过特征拼接的方式实现, 即: 将浅层特征图与深层特征图通过拼接的方式连接在一起。特征拼接在现有的许多网络模型中都有使用, 具体的, GoogLeNet 所使用的 Inception 模块^[39, 40]将不同尺度下的特征图拼接起来, UNet^[42]和 RED-Net^[43]将同一尺度的编码特征图和解码特征图拼接起来。通过简单的特征拼接可以有效地丰富所生成的特征, 而不必增加网络模型的参数规模。Huang 等^[32]进一步将特征拼接的思想用于构建 DenseNet 网络构架: 将所有的低层特征图与其后所有的高层特征图进行拼接。

本文提出的 DPCANet 中所使用的“稠密连接”与 DenseNet^[32]中所使用的“稠密连接”有如下几点不同。首先, 稠密连接的范围不同。为了防止高层网络参数规模的指数级增长, DenseNet 中的稠密连接局限于单个的稠密块 (Dense Block) 中; 而本文所使用的稠密连接局限于同一类型的卷积块或编码模块 (详见 4.2 节) 中。其次, 稠密程度不同。DenseNet 中的稠密连接将所有低层特征图与其后所有的高层特征图进行拼接, 因此, 在一个 L 层的稠密块中将发生 $L(L+1)/2$ 次连接; 而本文所使用的稠密连接仅将低层特征指向其后紧邻的下一个卷积层和最终的特征拼接层, 也就是, 除了最后一个特征块, 每个特征图只有 2 个后向连接, 因此, 一个 L 层的卷积块中只有 $2L-1$ 次连接。容易看出, $L(L+1)/2 \geq 2L-1$, 且当 $L > 2$ 时, $L(L+1)/2 > 2L-1$ 。之所以出现这样的稠密连接差异主要是由于 DenseNet 所采用的稠密连接更适用于面向多通道特征图的通道相关式卷积, 而在 PCANet 中只使用了通道无关式卷积, 并没有用到通道方向上的特征融合, 因此无须将低层特征图与其后续的所有高层特征图进行拼接。另外, PCANet 通过卷积将蕴含在像素域中的丰富特征逐渐在各个主方向上分离出来, 形成更具有判别性的特征空间。然而, 如果将第 $l-1$ 次和第 l 次的主方向分解的结果混合起来进行第 $l+1$ 次主方向分解, 就会将已经分离的特征再次混合起来 (此时所用到的主方向

会与第 l 次分解所用到的主方向混叠在一起), 这样就不利于特征的充分分解。

最后, 主要作用不同。正如 Huang 等^[32]所指出的, 在 DenseNet 中引入稠密连接的作用主要有 3 个: *i)* 可以极大地减少网络的参数, *ii)* 特征重用, *iii)* 强化特征在各网络层之间的传递, 从而可以有效地消除梯度消失问题。本文在 PCANet 中引入稠密连接主要是为了充分利用低层卷积层的特征图以增强特征的遮挡鲁棒性和多样性。

3.2 特征图的加权稠密编码

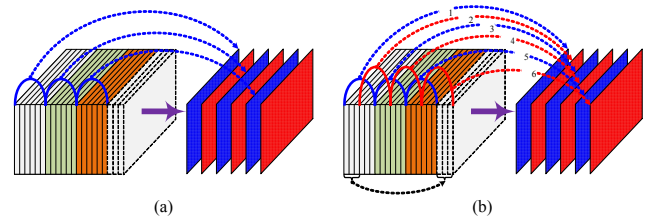


图 6 特征图的立体式稀疏编码(a)和加权立体式稠密编码(b)

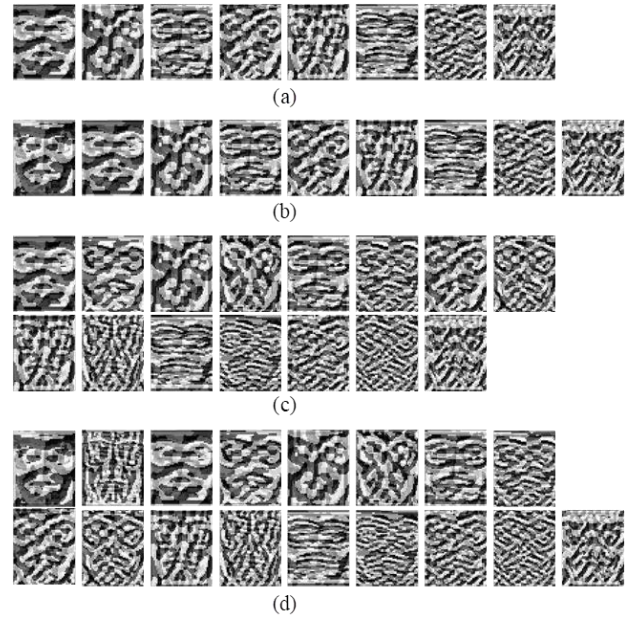


图 7 四种网络所生成的模式图: (a) PCANet, (b) DPCANet-I, (c) DPCANet-II, (d) DPCANet-III。这里, DPCANet-I 表示 DPCANet 只采用了对特征图的稠密连接, DPCANet-II 表示 DPCANet 只采用了对模式图的稠密编码, DPCANet-III 表示 DPCANet 同时采用了对特征图的稠密连接和对模式图的稠密编码。

PCANet 通过量化编码对最后一个卷积层的输出进行特征压缩, 这样做在去除冗余特征的同时也丢掉了大量的有用信息。为了解决上述问题, 本文引入稠密编码。图 6 给出了

PCANet 的立体式稀疏编码 (Stereo Sparse Coding, SSC) 和本文提出的立体式加权稠密编码 (Stereo Weighted Dense Coding, SWDC) 的示意图。SSC 采用稀疏步长：参与编码的各特征图之间没有重叠，而 SWDC 采用稠密步长：参与编码的各特征图之间有重叠。另外，我们也注意到，不同的模式图的遮挡鲁棒性是不同的，在 SWDC 中对不同的模式图施加了不同的权重。

就模式图的编码效果而言，本文提出的 SWDC 相对于 SSC 有两个优势：（1）基于 SWDC 可以获得比 SSC 更为丰富的模式编码图；（2）SSC 的编码方式实际上是将特征图的编码局限于由同一个 PCA 滤波器生成的特征图中，而忽略了不同 PCA 滤波器所生成的特征图之间的相关性，SWDC 对此进行了补充。图 7(a)和(c)分别展示基于 SSC 和 SWDC 所生成的模式图，对比图 7(a)和(c)可更为清楚地看出 SWDC 相对于 SSC 的上述两大优势。

显然，将特征图的稠密连接和稠密编码结合起来可以更为有效地生成更为丰富的特征，图 7(d)给出了将两者相结合后生成的模式图。可以看出，所生成的模式图相较于原始 PCANet 的模式图（图 7(a)）具有更为丰富的纹理特征和方向特征，从而可以有效弥补人脸图像中因遮挡或光照变化所丢失的判别信息。

3.3 权重参数计算

xPCANet 的参数主要有：用于稠密编码的权重参数 $\{\alpha_i\}$ 以及 xPCANet / xFPH 的三个分支中的权重参数 $\{\omega_i\}$ 。显然， $\{\alpha_i\}$ 和 $\{\omega_i\}$ 因输入图像、光照变化和遮挡图像等的不同而不同，通过手工调整 $\{\alpha_i\}$ 和 $\{\omega_i\}$ 的值很难使其适用于各种情形，而通过建立损失函数优化 $\{\alpha_i\}$ 和 $\{\omega_i\}$ 的值也是不现实的：一方面，需要在训练阶段使用有遮挡人脸图像，这与本文所定义的问题（图 2）相冲突了；另一方面， $\{\alpha_i\}$ 和 $\{\omega_i\}$ 的数量有限，通过学习的方法固定它们的值就很难使其自适应于不同的输入图像。本文提出基于柱状图特征稀疏性的方法对 $\{\alpha_i\}$ 和 $\{\omega_i\}$ 的值进行自动调整。

首先，值得注意的是，尽管两种权重参数 $\{\alpha_i\}$ 和 $\{\omega_i\}$ 的应用场景不同，它们在本质上是相同的：权重参数 $\{\omega_i\}$ 作用于 xPCANet 的各分支输出的柱状图特征 $\{H_i\}$ ，用于稠

密编码的权重 $\{\alpha_i\}$ 实际上也可视为作用于由各模式图输出的柱状图，因此，权重参数 $\{\alpha_i\}$ 和 $\{\omega_i\}$ 本质上都是用于平衡各柱状图特征的。为清晰起见，可以将 xPCANet 的输出通过如下公式展开：

$$f_{\text{xPCANet}}(\mathbf{x}) = \left[\omega_1 \begin{bmatrix} \alpha_{11} H_{11} \\ \alpha_{12} H_{12} \\ \vdots \\ \alpha_{1h_1} H_{1h_1} \end{bmatrix}^T, \omega_2 \begin{bmatrix} \alpha_{21} H_{21} \\ \alpha_{22} H_{22} \\ \vdots \\ \alpha_{2h_2} H_{2h_2} \end{bmatrix}^T, \omega_3 \begin{bmatrix} \alpha_{31} H_{31} \\ \alpha_{32} H_{32} \\ \vdots \\ \alpha_{3h_3} H_{3h_3} \end{bmatrix}^T \right]^T$$

其中， h_i 表示 xPCANet 的第 $i \in \{1, 2, 3\}$ 个分支所产生的模式图的个数， H_{ij} 表示由第 i 个分支的第 $j \in \{1, 2, \dots, h_i\}$ 个模式图所产生的柱状图特征。

其次，我们注意到，随着卷积层深度的增加，由各卷积层所输出的特征图所产生的柱状图的稀疏性也在逐渐降低。这是因为随着卷积层深度的增加，感受野在增大，各特征点融入了更多的邻域信息，在提取局部柱状图特征时，各滑动窗口内的特征也在逐渐变得丰富，所提取的柱状图特征的稀疏性也就降低了。对于有遮挡人脸识别而言，特征的稀疏性尤为重要，基于这一先验知识，可以对稀疏性较高的柱状图特征赋以较高的权值，对稀疏性较低的柱状图特征赋以较低的权值。柱状图特征 H 的稀疏性可以用如下公式计算：

$$f_{\text{sparse}}(H) = \|\mathbf{H} < \tau\|_1$$

其中， τ 为给定的阈值， $\|\cdot\|_1$ 表示 l_1 范数。 τ 的值与滑动窗口的大小有关，令滑动窗口的大小为 $w \times h$ ，本文设置 $\tau = wh/3$ ，即将一个滑动窗口中占比小于 1/3 的统计量设置为 0。

综上，可以通过如下方式计算 α_{ij} 和 ω_i ：

$$\alpha_{ij} = \frac{f_{\text{sparse}}(H_{ij})}{\sum_{k=1}^{h_i} f_{\text{sparse}}(H_{ik})}$$

$$\omega_i = \frac{f_{\text{sparse}}(H_i)}{\sum_{k=1}^3 f_{\text{sparse}}(H_k)}$$

4 实验

4.1 实验设置

为了验证本文方法的有效性和通用性，针对受控和非受控两种环境下的人脸图像数据集进行了测试。在受控环境下，主要使用了 AR 人脸数据集^[44]。AR 数据集是经典的人脸数据集，包含了强光光照、太阳镜、围巾等三种类型的自然遮挡，详见 5.3 节；在非受控环境下，采用了与 Zhang 等人的

工作^[45]类似的人脸数据集, 主要包括 LFW^[46]、CFP^[47]、MFR2^[48]和 PKU-Masked-Face^[33]等四个数据集。LFW 和 CFP 所包含的有遮挡人脸图像较少, 我们使用人脸遮挡生成工具^[48]来生成模拟遮挡, 详见 5.4 节; MFR2 和 PKU-Masked-Face 包含了大量的有实际遮挡的人脸图像, 详见 5.5 节。另外, 为了对 xPCANet 模型进行训练, 选择 WebFace^[49]作为训练集, WebFace 是一个大规模的人脸数据集, 由来自 10,575 个人的 494,414 张人脸图像组成, 并且 WebFace 中大部分人脸图像不含遮挡, 符合本文在第二节中对有遮挡人脸识别问题的定义。

为了评估本文方法的有效性, 将 xPCANet 框架与不同的 CNN 模型相结合, 对比不同的 CNN 模型在使用 xPCANet 框架前后的遮挡鲁棒性的变化。针对受控环境数据集和非受控环境数据集采用了不同的 CNN 模型。对于受控环境数据集, 选取传统的机器学习方法 NMR^[50]以及两个轻量级的 CNN 模型: VGGFace^[15]和 LCNN (Lightened CNN)^[51]; 对于非受控环境数据集, 选取 VGGFace^[15]、FaceNet^[2]和 SphereFace2^[17, 52]等三种针对一般人脸图像 (不含遮挡) 进行训练的 CNN 模型, 以及 InterpretFR^[34]、LPD^[33]和 ECBAM^[45]等三种针对有遮挡人脸图像进行训练的 CNN 模型。VGGFace、FaceNet、InterpretFR 和 LPD 的作者提供了公开可获取的预训练模型³; LCNN、SphereFace2、ECBAM 的作者提供了训练网络模型的源码⁴, 本文统一使用 WebFace 对其进行训练。为简单起见, 用下划线标注使用了 xPCANet 框架的 CNN 模型, 例如, LCNN 表示使用了 xPCANet 框架的 LCNN 模型。

xPCANet 基于现有的 CNN 模型和 DPCANet。本文将现有的 CNN 模型视作黑盒子, 不对其进行配置和更改, 主要验证 xPCANet 在不同的稠密连接和权重配置下的性能。5.2 节将针对 DPCANet 在不同情形下的配置进行消融实验。5.3~5.5 节针对不同的数据集和不同的 CNN 模型验证

xPCANet 的有效性。根据 Chan 等人的工作^[29], 在各种情形下 xPCANet 的特征提取部分均采用两层卷积, 且各卷积层所采用的卷积核个数均为 8。

对不同类型的特征采用不同的分类器: 由于 xPCANet 最终生成的是柱状图特征, 采用基于卡方距离^[29]的最近邻分类器 (详见公式(1)); 对 NMR、VGG 和 LCNN 统一采用基于协同表示的分类器^[53]。

在 Win10 64 位操作系统下部署实验, 主要硬件配置为: 8G 内存 (RAM), Intel i7-4600 双核处理器 (主频分别为 2.1GHz 和 2.69GHz)。所有方法均使用 Python 语言实现, 并使用 PyTorch 作为深度学习框架。

4.2 消融实验

从稠密连接的角度来看, 需要考虑 xPCANet 的三种配置: 只对特征图进行稠密连接 (用 DF 表示)、只对模式图进行稠密编码 (用 DP 表示)、同时使用 DF 和 DP; 从是否使用权重的角度来看, 也需要考虑 xPCANet 的三种配置: 只在稠密编码时使用权重 $\{\alpha_i\}$ (用 α 表示)、只对 xPCANet 三个网络分支的输出使用权重 $\{\omega_i\}$ (用 ω 表示)、同时使用 α 和 ω 。选取 VGGFace 作为 xPCANet 内嵌的 CNN 模型, 选取 AR (Scarf)、LFW-Occ (U-O)、MFR2 (U-O)、PKU-Masked-Face (U-O) 等数据子集作为测试集, 验证 xPCANet 在 DF、DP、 α 和 ω 等四种配置及它们的组合下的识别性能。各测试子集的详细描述见 5.3 节~5.5 节。

表 1 xPCANet 的消融实验

DF	DP	α	ω	AR (Scarf)	LFW-Occ (U-O)	MFR2 (U-O)	PKU-Masked-Face (U-O)
				89.69	93.46	90.25	87.25
✓				91.72	94.78	91.29	88.75
	✓			90.15	94.59	90.58	88.69
✓	✓			92.36	95.57	91.57	89.39
		✓		90.28	94.29	90.63	87.18
			✓	90.43	93.69	90.37	87.83
		✓	✓	91.86	95.29	91.32	88.59
✓	✓	✓	✓	93.28	97.15	91.63	90.00

表 1 给出了消融实验的结果。从第 2 行~第 3 行中呈现的结果来看, 一般情况下, DF 比 DP 更为重要, 能更好地提升识别性能。由于稠密连接使用了网络的底层特征, 这说明对于有遮挡人脸识别而言, 网络的底层特征比网络的高层

³ VGGFace: https://www.robots.ox.ac.uk/~vgg/software/vgg_face/

FaceNet: <https://github.com/timesler/facenet-pytorch>

InterpretFR: <http://cvlab.cse.msu.edu/project-interpret-FR>

LPD

<https://gitee.com/ffeiding/Masked-Face-Recognition-with-Latent-Part-Detection>

⁴ LCNN: https://github.com/AlfredXiangWu/face_verification_experiment

SphereFace2: <https://opensphere.world/>

ECBAM: <https://github.com/Jongchan/attention-module>

特征更为重要。比较表 1 的第 2 行~第 4 行中的结果可以看出, DF+DP 能获得比单独使用 DF 或 DP 更好的性能。从第 5 行~第 6 行中的结果可以看出, 一般情况下, α 比 ω 能更好地提升识别性能, 这说明调整各网络分支内部的各柱状图特征的权重比调整各分支输出的柱状图特征的权重更为重要。比较第 5 行~第 7 行的结果可以看出, 联合使用 α 和 ω 能获得比单独使用 α 或 ω 更好的性能。表 1 的最后一行给出了同时使用 DF、DP、 α 和 ω 的 xPCANet 的识别性能, 可以看出, 此时 xPCANet 的识别性能最高。

4.3 受控环境下实际遮挡数据上的识别性能

本节使用受控环境下采集的 AR 人脸图像数据集^[44] 测试 xPCANet 在有实际遮挡和极端光照变化情形的识别性能。从 AR 库中选取 119 个人 (65 个男性和 54 个女性) 的图像作为实验数据。训练集选取 $119 \times 8 = 952$ 张不含遮挡和光照变化的人脸图像, 如图 (d) 所示; 测试集 I 选取 $119 \times 3 = 357$ 张有左侧、右侧和正面等三种强光变化的人脸图像, 如图 8(a) 所示; 测试集 II 选取 357 张有太阳镜遮挡且分别包含了三种强光变化的人脸图像, 如图 8(b) 所示; 测试集 III 选取 357 张有围巾遮挡且混合了三种强光变化的人脸图像, 如图 8(c) 所示。AR 库中的强光光照与遮挡类似, 造成了大面积的判别特征的损失, 具有更大的识别难度。所有图像首先被裁剪和对齐为 56×46 的像素矩阵。为了测试各方法的识别性能随着训练图像分辨率的变化情况, 对所有训练/测试图像进行了 4 倍、3 倍、2 倍和 1 倍下采样, 分别对应于 $14 \times 11=154$ 、 $18 \times 15=270$ 、 $28 \times 23=644$ 和 $56 \times 46=2576$ 等像素维数。

表 2、表 3 和表 4 分别针对图 8 所给出的训练集和三个测试子集, 对比了 xPCANet 相对于经典的机器学习方法 NMR^[50] 和两个轻量级的 CNN 模型 VGGFace^[15] 和 LCNN (Lightened CNN)^[51] 的有效性。可以看出: VGGFace 相对于 VGGFace、LCNN 相对于 LCNN, 在各个测试集上都取得了显著的性能提升, 尤其是当测试集 (测试 II 和测试集 III) 中有遮挡存在时; 并且, 在大多情况下, VGGFace 和 LCNN 的识别性能都超过了原始的 PCANet。

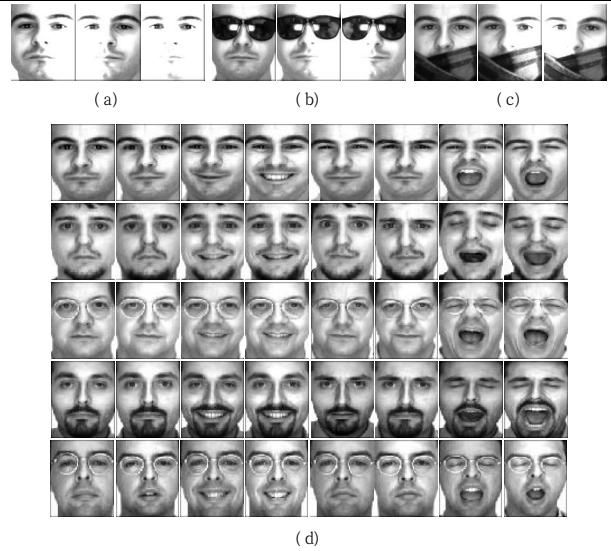


图 8 来自 AR 人脸数据库的训练集样本和测试集样本, 其中, (a)~(c) 为测试集 I~III 的样本样例, (d) 为训练集的样本样例

表 2 在 AR 人脸库的测试集 I 上的识别率 (%)

像素维	154	270	644	2576
NMR	48.50	70.31	77.19	83.51
PCANet	44.82	68.35	75.91	82.07
VGGFace	43.25	60.32	60.58	69.85
<u>VGGFace</u>	47.34	69.19	77.47	85.35
LCNN	10.36	38.53	46.52	60.25
<u>LCNN</u>	45.66	69.19	75.93	79.83

表 3 在 AR 人脸库的测试集 II 上的识别率 (%)

像素维	154	270	644	2576
NMR	23.41	50.42	73.99	76.20
PCANet	19.61	46.22	70.03	89.64
VGGFace	5.32	9.80	18.72	22.69
<u>VGGFace</u>	22.69	48.18	72.55	90.20
LCNN	21.57	43.70	46.79	59.66
<u>LCNN</u>	21.62	46.50	71.15	90.20

表 4 在 AR 人脸库的测试集 III 上的识别率 (%)

像素维	154	270	644	2576
NMR	33.49	60.78	78.15	89.90
PCANet	31.37	57.14	74.79	91.88
VGGFace	9.52	54.90	58.90	68.63
<u>VGGFace</u>	31.37	59.10	78.87	93.28
LCNN	42.30	67.51	70.58	71.99
<u>LCNN</u>	43.49	69.94	77.31	92.72

与其它方法相比, VGGFace 不管在高像素维还是低像素维中, 它的识别率都是最低的。这说明 VGGFace 对光照和遮挡的鲁棒性并不是很好, 这很大程度上与网络结构的设

计及其预训练模型所使用的训练数据集有关。与 FaceNet 的预训练模型一样, VGGFace 使用 VGGFace2^[54]作为训练集, 而 VGGFace2 主要关心的是姿态和年龄的变化。但在 xPCANet 框架下, VGGFace 的遮挡鲁棒性和光照鲁棒性均有了显著提升。

与其它方法相比, LCNN 在待识别图像的分辨率较低的时候, 性能更好, 这一点在测试集 III 上的体现最为明显: 当像素维分别为 154 和 270 时, LCNN / LCNN 在测试集 III 上的识别率分别达到了 42.30% / 43.49% 和 67.51% / 69.94%, 超过了其它所有方法。这是因为 LCNN 对于低分辨率图像的特征提取能力比其它方法要更好些, 但是随着图像分辨率的提高, xPCANet 通过稠密连接后, 能够提取到更丰富的特征, VGGFace 取得了更好的性能。如何在低分辨率情境下进一步提升 xPCANet 的识别性能有待于进一步讨论。

4.4 非受控环境下模拟遮挡数据上的识别性能比较



图9 本节实验所用到的非受控环境下的模拟遮挡人脸数据集的样本示例

本节及下一节评估本文方法在非受控环境下的识别性能, 本节重点关注各对比法对非受控环境下的模拟遮挡的处理能力。在两个常用的非受控环境下的基准人脸数据集 LFW^[46]和 CFP^[47]上人工合成有遮挡的人脸图像数据集。CFP 数据集由两部分组成: 第一部分仅由正面人脸(Frontal Face, FF) 图像组成, 称为 CFP-FF; 第二部分混合了正面和非正面人脸图像, 称为 CFP-FP (Frontal Profile, FP)。使用人脸遮挡生成工具^[48]来模拟遮挡数据集, 选择具有不同颜色和纹理的口罩遮挡添加到 LFW、CFP-FF 和 CFP-FP 上, 分别得到新的有遮挡的人脸数据集: LFW-Occ、CFP-FF-Occ 和 CFP-FP-Occ。这里, 在 CFP-FP 子集上, 只对其正面脸图像

添加遮挡, 而非正面脸图像则保持原状, 未添加遮挡。这一方面是因为非正面脸图像已经形成了“自遮挡”, 为简化问题, 就不再为其添加遮挡; 另一方面则是因为我们使用的遮挡合成工具^[48]的性能在很大程度上依赖于对人脸图像中基准位置的检测, 而这些基准位置在非正面脸图像中不容易检测到。图9展示了本节实验所用到的非受控环境下的有遮挡人脸数据集中的样本。

表5 在三个模拟遮挡人脸数据集上的最大识别率 (%)

Models	LFW-Occ			CFP-FF-Occ			CFP-FP-Occ		
	U-U	U-O	O-O	U-U	U-O	O-O	U-U	U-O	O-O
VGGFace	97.73	93.46	94.46	<u>97.74</u>	89.74	90.87	78.52	73.12	74.38
<u>VGGFace</u>	97.68	97.15	97.06	<u>96.04</u>	95.20	94.65	<u>77.48</u>	77.18	76.59
FaceNet	93.28	94.39	93.25	92.38	90.20	90.15	78.57	77.35	76.71
<u>FaceNet</u>	94.51	95.56	94.35	93.41	92.33	91.86	79.65	80.23	80.15
SphereFace2	97.95	97.35	97.22	95.38	94.77	94.43	79.43	78.16	77.33
<u>SphereFace2</u>	97.89	98.59	<u>97.38</u>	95.47	95.58	94.52	80.53	80.23	79.42
InterpretFR	97.66	93.38	93.03	94.00	91.54	91.31	80.35	78.51	79.38
<u>InterpretFR</u>	<u>98.32</u>	97.53	94.25	94.52	<u>96.21</u>	<u>96.15</u>	80.72	80.71	80.05
LPD	94.82	94.26	93.70	90.48	89.50	89.00	80.92	80.38	79.61
<u>LPD</u>	<u>94.80</u>	95.83	94.62	90.50	93.27	91.46	<u>81.09</u>	<u>81.16</u>	<u>80.21</u>
ECBAM	98.02	97.60	97.33	96.30	95.96	95.53	80.65	79.13	78.87
<u>ECBAM</u>	98.71	<u>98.30</u>	98.35	<u>97.42</u>	96.32	96.37	81.37	81.19	81.75

在 1:1 情境下比较各方法的人脸验证 (Face Verification) 的性能。具体的, 在如下三种情形下进行测试: (1) 不含遮挡 vs. 不含遮挡 (U-U), 即一般的人脸验证任务, 相比对的各人脸图像对都不含遮挡, 用于测试模型的通用性; (2) 不含遮挡 vs. 包含遮挡 (U-O), 即: 相比对的各人脸图像对中有一幅不含遮挡, 另一幅含有遮挡; (3) 遮挡 vs. 遮挡 (O-O), 相比对的各人脸图像对均含有遮挡。为了与现有方法进行公平对比, 进行了 10 折交叉验证, 并选取各方法的最高识别率来评估该方法的性能。

表5 报告了相关的实验结果。为清晰起见, 表中对最优识别率和次优识别率分别进行了**加粗**和**加下划线**标注, 对每个模型在每个测试子集的三种情形 (U-U、U-O、O-O) 下的最低识别率进行了单元格加阴影标注, 对施加了 xPCANet 却出现性能退化的模型进行了“**斜体+字体阴影**”标注。可以看出, 使用 xPCANet 在绝大多数情境下都能够有效提升

原有 CNN 模型的识别性能, 只在 U-U 的情形下导致了少部分 CNN 模型 (VGGFace、SphereFace2、LPD) 的性能退化, 这是因为 xPCANet 主要是用于解决有偏移的图像识别问题, 而在 U-U 情形下, 待识别图像和库存图像并没有发生显著的偏移, xPCANet 的优越性也很难体现。

在 LFW-Occ 和 CFP-FF-Occ 两个测试集上, VGGFace、FaceNet 和 SphereFace2 等三种针对一般人脸图像 (不含遮挡) 进行训练的模型, 取得了可以与 InterpretFR、LPD 和 ECBAM 等三种针对有遮挡人脸图像进行训练的模型相比较的性能, 而在更有挑战的数据集 CFP-FP-Occ 上, VGGFace、FaceNet 和 SphereFace2 的表现明显弱于 InterpretFR、LPD 和 ECBAM。这一方面是因为 VGGFace 和 FaceNet 等的预训练模型是基于 VGGFace2^[54]进行训练的, VGGFace2 更强调对姿势变化的处理, 而姿势变化会引发“自遮挡”, 在某种程度上扮演了“遮挡”的角色; 另一方面, 正如 Sun 等^[55]所指出的, 深度神经网络模型本身具有一定的遮挡鲁棒性, 即使不针对“遮挡”进行设计, 对于一般的遮挡问题, 也能达到一定的鲁棒性。然而, 当遮挡较为严重时, 如对于 CFP-FP-Occ 测试集, 对遮挡进行专门的建模和处理就显得非常必要了。

另外, U-U、U-O 和 O-O 这三种情形下, 哪种情形最有挑战? U-O 是本文第 2 节定义的问题, 由于无遮挡的库存图像和有遮挡的测试图像之间存在着显著的“偏移”, 直觉上可能会认为 U-O 更有挑战性。然而, 表 5 的实验结果表明, 在大多数情境下, 反而是 O-O 更有挑战, 这是因为当相比对的两幅图像同时包含遮挡时, 遮挡在两幅图像中的位置在很大概率下是不同的, 这就意味着, 两幅图像中的不同位置的判别性特征同时都被挡住了, 从而导致分类器发生误判。

值得注意的是, 在表 5 中并没有报告两个前沿方法 PDSN^[13]和 DFM^[56]的识别率, 这是因为 PDSN 和 DFM 没有公开可获取的预训练模型, 而且它们的网络模型非常复杂, 本文实验所用到的硬件环境很难进行重新进行训练。这里直接引用 Ding 等人^[33]报告的在 LFW 上的结果来评估 PDSN 和 DFM 的识别性能: 二者在 LFW-Occ 的 U-O 测试环境下的识别率分别为 86.72%和 92.88%。对比表 5 中所报告的其

它方法的识别性能, 可以看出 PDSN 和 DFM 相对于本文所使用的 CNN 模型的识别性能还是有一定差距的。这表明过于复杂的方法的通用性往往也有一定的局限性。

4.5 非受控环境下真实遮挡数据上的识别性能比较

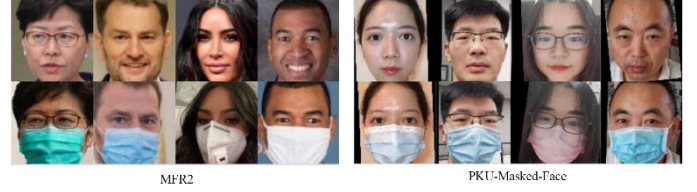


图 10 本节实验所用到的非受控环境下的实际遮挡人脸数据集的样本示例

本节评估本文方法在非受控环境下面向真实遮挡时的识别性能。为此, 我们在最近建立的两个有真实遮挡的人脸数据集 MFR2 和 PKU-Masked-Face 上进行实验。MFR2^[48]是一个小型数据集, 共有来自 53 个人的 269 张人脸图像; 而 PKU-Masked-Face^[33]是一个大规模的数据集, 它由来自 1018 个不同人的 10301 张面部图像组成, 其中每个人都包含了一定数量的不含任何遮挡的面部图像和含有遮挡的面部图像, 并且在姿势、光照和遮挡类型等方面都有较大的变化。为了重点测试“遮挡”的影响, 我们只选择正面脸的有遮挡和无遮挡图像进行测试: 在 MFR2 数据集上, 随机选取 800 个图像对 (400 个正样本对和 400 个负样本对); 在 PKU-Masked-Face 人脸数据集上, 随机选取 1890 个图像对 (945 个正样本对和 945 个负样本对)。

表 6 报告了各对比方法在两个有真实遮挡人脸数据集上的最大识别率。显然, 使用 xPCANet 在绝大多数情境下都能够有效提升原有 CNN 模型的识别性能, 只在无偏移的 U-U 情形下导致了 VGGFace 和 FaceNet 的性能退化。另外, 值得注意的是, 从 MFR2 的三个测试子集的测试结果来看, 对于 VGGFace、FaceNet 和 SphereFace2 等三种针对一般人脸图像 (不含遮挡) 进行训练的模型, O-O 情形仍然是最有挑战的; 然而, 对于 InterpretFR、LPD 和 ECBAM 等三种针对有遮挡人脸图像进行训练的模型, U-O 情形却是最有挑战的。导致这种差异的一个重要原因在于 MFR2 数据集主要包含的是口罩遮挡, 而口罩在人脸表面所覆盖的区域大致

接近，因此，能够有效消除遮挡影响的遮挡鲁棒模型（如 InterpretFR 等）可以取得更优的识别性能，在此情形下，经典的 CNN 模型（如 VGGFace 等）由于无法较好地消除遮挡的影响，仅凭提取遮挡鲁棒特征，所生成的判别性特征仍然受到了很大影响，其识别性能弱于 U-U 和 U-O 两种情形。在 PKU-Masked-Face 数据集上只测试了本文重点关注的 U-O 情境下的各模型的识别性能。可以看出，xPCANet 在大规模有实际遮挡的非受控环境下也可以稳定提升现有网络模型的识别性能。

表 6 在两个真实遮挡人脸数据集上的最大识别率 (%)

Models	MFR2		PKU-Masked-Face	
	U-U	U-O	O-O	U-O
VGGFace	99.85	90.14	86.74	88.37
<u>VGGFace</u>	99.78	91.63	87.62	90.00
FaceNet	99.67	92.01	91.50	89.58
<u>FaceNet</u>	99.59	93.27	93.67	91.67
SphereFace2	99.79	93.38	92.65	90.75
<u>SphereFace2</u>	99.80	93.79	93.55	92.59
InterpretFR	98.17	90.63	91.75	92.85
<u>InterpretFR</u>	<u>99.90</u>	92.42	<u>94.51</u>	<u>94.15</u>
LPD	97.20	88.76	91.86	91.96
<u>LPD</u>	98.52	90.54	92.73	92.85
ECBAM	99.92	92.38	93.37	93.97
<u>ECBAM</u>	99.92	<u>93.56</u>	95.76	94.52

5 结论

PCANet 对强光光照变化、遮挡等具有良好的鲁棒性，然而，PCANet 主要适用于受控环境下的人脸识别；现有的 CNN 模型能够适用于非受控环境下的人脸识别，然而，往往缺乏遮挡鲁棒性。为了在非受控环境下达到较好的遮挡鲁棒性，本文将 PCANet 与现有的 CNN 模型相结合，提出了扩展的 PCANet 模型：xPCANet。为了使 PCANet 能够更好地适用于 CNN 模型，对 PCANet 模型进行了改进。PCANet 的主要不足在于：所使用的滤波器缺乏多样性，未能充分利用网络的底层卷积层特征，在模式图编码的过程中丢掉了过多的信息。而滤波器多样性和滤波器响应的多样性对于有遮挡人脸图像的识别而言尤为重要。通过对 PCANet 的卷积层特征进行稠密连接、对 PCANet 的模式图进行加权稠密编码，可以有效解决上述问题。在受控环境和非受控环境下的实验

均表明了 xPCANet 的有效性。由于消除遮挡的影响对于提升遮挡鲁棒性具有重要意义（详见本文 5.5 节），未来将针对这一问题进一步提出解决方案。

参考文献

- [1] TAIGMAN Y, YANG M, RANZATO M A, et al. DeepFace: Closing the Gap to Human-Level Performance in Face Verification [C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New York: IEEE Press, 2014: 1701-1708.
- [2] SCHROFF F, KALENICHENKO D, PHILBIN J. FaceNet: A Unified Embedding for Face Recognition and Clustering [C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New York: IEEE Press, 2015: 815-823.
- [3] HE K, ZHANG X, REN S, et al. Deep Residual Learning for Image Recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 770-778.
- [4] HE K, ZHANG X, REN S, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on Imagenet Classification [C]// Proceedings of the International conference on computer vision. 2015: 1026-1034.
- [5] IOFFE S, SZEGEDY C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift [C]// Proceedings of the International conference on machine learning. 2015: 448-456.
- [6] QUONERO-CANDELA J, SUGIYAMA M, SCHWAIGHOFER A, et al. Dataset Shift in Machine Learning [M]. The MIT Press, 2009.
- [7] SHIMODAIRA H. Improving Predictive Inference under Covariate Shift by Weighting the Log-Likelihood Function [J]. Journal of statistical planning and inference, 2000, 90(2): 227-244.
- [8] ZHAO F, FENG J, ZHAO J, et al. Robust Lstm-Autoencoders for Face De-Occlusion in the Wild [J]. IEEE Transactions on Image Processing, 2017, 27(2): 778-790.
- [9] WRIGHT J, YANG A, GANESH A, et al. Robust Face Recognition Via Sparse Representation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.
- [10] ZHOU Z, WAGNER A, MOBAHI H, et al. Face Recognition with Contiguous Occlusion Using Markov Random Fields [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2009: 1050-1057.
- [11] LI X-X, DAI D-Q, ZHANG X-F, et al. Structured Sparse Error Coding for Face Recognition with Occlusion [J]. IEEE Transactions on Image Processing, 2013, 22(5): 1889-1900.
- [12] WAN W, CHEN J. Occlusion Robust Face Recognition Based on Mask Learning [C]// Proceedings of the IEEE/CVF International Conference on Image Processing.

- New York: IEEE Press, 2017: 3795-3799.
- [13] SONG L, GONG D, LI Z, et al. Occlusion Robust Face Recognition Based on Mask Learning with Pairwise Differential Siamese Network [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2019: 773-782.
 - [14] QIU H, GONG D, LI Z, et al. End2end Occluded Face Recognition by Masking Corrupted Features [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 44(10): 6939-6952.
 - [15] PARKHI O M, VEDALDI A, ZISSERMAN A. Deep Face Recognition [C]// Proceedings of the British Machine Vision Conference. Durham: BMVA Press, 2015: 6.
 - [16] HE K, ZHANG X, REN S, et al. Identity Mappings in Deep Residual Networks [C]// Proceedings of the European Conference on Computer Vision. Springer, 2016: 630-645.
 - [17] LIU W, WEN Y, RAJ B, et al. Sphereface Revived: Unifying Hyperspherical Face Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 45(2): 2458-2474.
 - [18] TZIMIROPOULOS G, ZAFEIRIOU S, PANTIC M. Subspace Learning from Image Gradient Orientations [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2012, 34(12): 2454-2466.
 - [19] LI X-X, HAO P, HE L, et al. Image Gradient Orientations Embedded Structural Error Coding for Face Recognition with Occlusion [J]. Journal of Ambient Intelligence and Humanized Computing, 2020, 11(6): 2349-2367.
 - [20] ZHANG W, SHAN S, GAO W, et al. Local Gabor Binary Pattern Histogram Sequence (Lgbphs): A Novel Non-Statistical Model for Face Representation and Recognition [C]// Proceedings of the IEEE International Conference on Computer Vision. IEEE, 2005: 786-791.
 - [21] AHONEN T, HADID A, PIETIKAINEN M. Face Description with Local Binary Patterns: Application to Face Recognition [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2006, 28(12): 2037-2041.
 - [22] SZEGEDY C, ZAREMBA W, SUTSKEVER I, et al. Intriguing Properties of Neural Networks [C]// Proceedings of the International Conference on Learning Representations. 2014.
 - [23] NGOC-SON V, CAPLIER A. Enhanced Patterns of Oriented Edge Magnitudes for Face Recognition and Image Matching [J]. IEEE Transactions on Image Processing, 2012, 21(3): 1352-1365.
 - [24] NGOC-SON V. Exploring Patterns of Gradient Orientations and Magnitudes for Face Recognition [J]. IEEE Transactions on Information Forensics and Security, 2013, 8(2): 295-304.
 - [25] AHONEN T, HADID A, PIETIKAINEN M. Face Recognition with Local Binary Patterns [C]// Proceedings of the European Conference on Computer Vision. 2004: 469-481.
 - [26] ZHANG B, GAO Y, ZHAO S, et al. Local Derivative Pattern Versus Local Binary Pattern: Face Recognition with High-Order Local Pattern Descriptor [J]. IEEE Transactions on Image Processing, 2009, 19(2): 533-544.
 - [27] YANG M, ZHANG L, SHIU S C-K, et al. Monogenic Binary Coding: An Efficient Local Feature Extraction Approach to Face Recognition [J]. IEEE Transactions on Information Forensics and Security, 2012, 7(6): 1738-1751.
 - [28] YANG M, ZHANG L, SHIU S C-K, et al. Robust Kernel Representation with Statistical Local Features for Face Recognition [J]. IEEE Transactions on Neural Networks And Learning Systems, 2013, 24(6): 900-912.
 - [29] CHAN T-H, JIA K, GAO S, et al. PCANet: A Simple Deep Learning Baseline for Image Classification? [J]. IEEE Transactions on Image Processing, 2015, 24(12): 5017-5032.
 - [30] 李小薪, 吴克宋, 齐盼盼, et al. 局部球面规范化嵌入: PCANet 的一种改进方案 [J]. 计算机科学, 2018, 45(5): 238-242.
 - [31] LOW C-Y, TEOH A B-J, TOH K-A. Stacking PCANet+: An Overly Simplified Convnets Baseline for Face Recognition [J]. IEEE Signal Processing Letters, 2017, 24(11): 1581-1585.
 - [32] HUANG G, LIU Z, VAN DER MAATEN L, et al. Densely Connected Convolutional Networks [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4700-4708.
 - [33] DING F, PENG P, HUANG Y, et al. Masked Face Recognition with Latent Part Detection [C]// Proceedings of the ACM international Conference on multimedia. 2020: 2281-2289.
 - [34] YIN B, TRAN L, LI H, et al. Towards Interpretable Face Recognition [C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. New York: IEEE Press, 2019: 9348-9357.
 - [35] 李小薪, 梁荣华. 有遮挡人脸识别综述: 从子空间回归到深度学习 [J]. 计算机学报, 2018, 41(1): 177-207.
 - [36] LOW C-Y, TEOH A B-J, NG C-J. Multi-Fold Gabor, Pca and Ica Filter Convolution Descriptor for Face Recognition [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2017:
 - [37] TAI Y, YANG J, LIU X, et al. Memnet: A Persistent Memory Network for Image Restoration [C]// Proceedings of the IEEE International Conference on Computer Vision. 2017: 4539-4547.
 - [38] CHEN D, CAO X, WEN F, et al. Blessing of Dimensionality: High-Dimensional Feature and Its Efficient Compression for Face Verification [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2013: 3025-3032.
 - [39] SZEGEDY C, LIU W, JIA Y, et al. Going Deeper with Convolutions [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 1-9.
 - [40] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the Inception Architecture for Computer Vision [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2818-2826.
 - [41] REN C-X, DAI D-Q, LI X-X, et al. Band-Reweighted Gabor Kernel Embedding for Face Image Representation and Recognition [J]. IEEE Transactions on Image

- Processing, 2013, 23(2): 725-740.
- [42] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [C]// Proceedings of the International Conference on Medical image computing and computer-assisted intervention. New York: Springer, 2015: 234-241.
- [43] MAO X, SHEN C, YANG Y-B. Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections [C]// Proceedings of the Advances in Neural Information Processing Systems. Cambridge: MIT Press, 2016: 2802-2810.
- [44] MARTINEZ A M, The AR Face Database: CVC, 1998.
- [45] ZHANG Y, WANG X, SHAKEEL M S, et al. Learning Upper Patch Attention Using Dual-Branch Training Strategy for Masked Face Recognition [J]. Pattern Recognition, 2022, 126: 108522.
- [46] HUANG G B, MATTAR M, BERG T, et al., Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments: 07-49 [R]. Amherst: University of Massachusetts, 2008.
- [47] SENGUPTA S, CHEN J-C, CASTILLO C, et al. Frontal to Profile Face Verification in the Wild [C]// Proceedings of the IEEE winter conference on applications of computer vision (WACV). IEEE, 2016: 1-9.
- [48] ANWAR A, RAYCHOWDHURY A, Masked Face Recognition for Secure Authentication: arXiv preprint arXiv:2008.11104 [R], 2020.
- [49] YI D, LEI Z, LIAO S, et al. Learning Face Representation from Scratch [J]. arXiv preprint arXiv:14117923, 2014: 1-10.
- [50] YANG J, LUO L, QIAN J, et al. Nuclear Norm Based Matrix Regression with Applications to Face Recognition with Occlusion and Illumination Changes [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(1): 156-171.
- [51] WU X, HE R, SUN Z. A Lightened CNN for Deep Face Representation [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
- [52] LIU W, WEN Y, YU Z, et al. Sphreface: Deep Hypersphere Embedding for Face Recognition [C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New York: IEEE Press, 2017: 212-220.
- [53] ZHANG L, YANG M, FENG X. Sparse Representation or Collaborative Representation: Which Helps Face Recognition? [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2011: 471-478.
- [54] CAO Q, SHEN L, XIE W, et al. Vggface2: A Dataset for Recognising Faces across Pose and Age [C]// Proceedings of the IEEE international conference on automatic face & gesture recognition. IEEE, 2018: 67-74.
- [55] SUN Y, WANG X, TANG X. Deeply Learned Face Representations Are Sparse, Selective, and Robust [C]// Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. New York: IEEE Press, 2015: 2892-2900.
- [56] HE L, LI H, ZHANG Q, et al. Dynamic Feature Learning for Partial Face Recognition [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7054-7063.

Occluded Face Recognition via Extended PCANet

QIN E*, LU Tian-Yu*, LI Weifeng**, LIU Yin-Wei*, ZHU Yani***, LI Xiao-Xin*
(* College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023)
(** Watone Cloud Data Co., Ltd, Hangzhou 310018)
(*** School of Cyberspace, Hanzhou Dianzi University, Hangzhou 310018)

Abstract

To solve the problem of occluded face recognition, an extended PCANet (xPCANet) model is proposed by combining the existing CNN model with PCANet. In order to effectively eliminate the influence caused by facial occlusions, it is usually necessary to make full use of the low-level features of the network and build rich features. Two disadvantages of PCANet are: due to the orthogonality constraint, the filters of each convolution layer are highly similar, which reduces the diversity of the filter response; to get the pattern maps, the feature maps are binarized, and the encoding method with a large stride is adopted, so that many useful features are discarded. In order to make PCANet better fit into the existing CNN models, we introduce two dense connections into the PCANet model: the dense connections introduced between convolutional layers are used to make full use of the features extracted by the low-level convolutional layers, and reduce the similarity of filters between convolutional layers as much as possible; in the pattern-map encoding stage, weighted dense encoding is introduced to make full use of the features produced by the convolutional layers to generate more pattern maps. These two dense connections enhance the dimension of the final output of PCANet histogram features and generate richer features. Experiments on face datasets (AR face dataset) with real occlusions acquired in the controlled environment, on face datasets (LFW and CFP) with synthetic occlusions acquired in the uncontrolled environment, and on face datasets (MFR2 and PKU-Masked Face) with real occlusions acquired in the uncontrolled environment show that, compared with existing methods, the proposed DPCANet can effectively deal with physical occlusions and illumination-caused occlusions, and can also be an effective supplement to the cutting-edge methods to improve their robustness against occlusions.

Key words: face recognition with occlusion, PCANet, dense connections, dense encoding, filter diversity
